Robust Logic Reasoning with Graph Neural Networks

Le Song

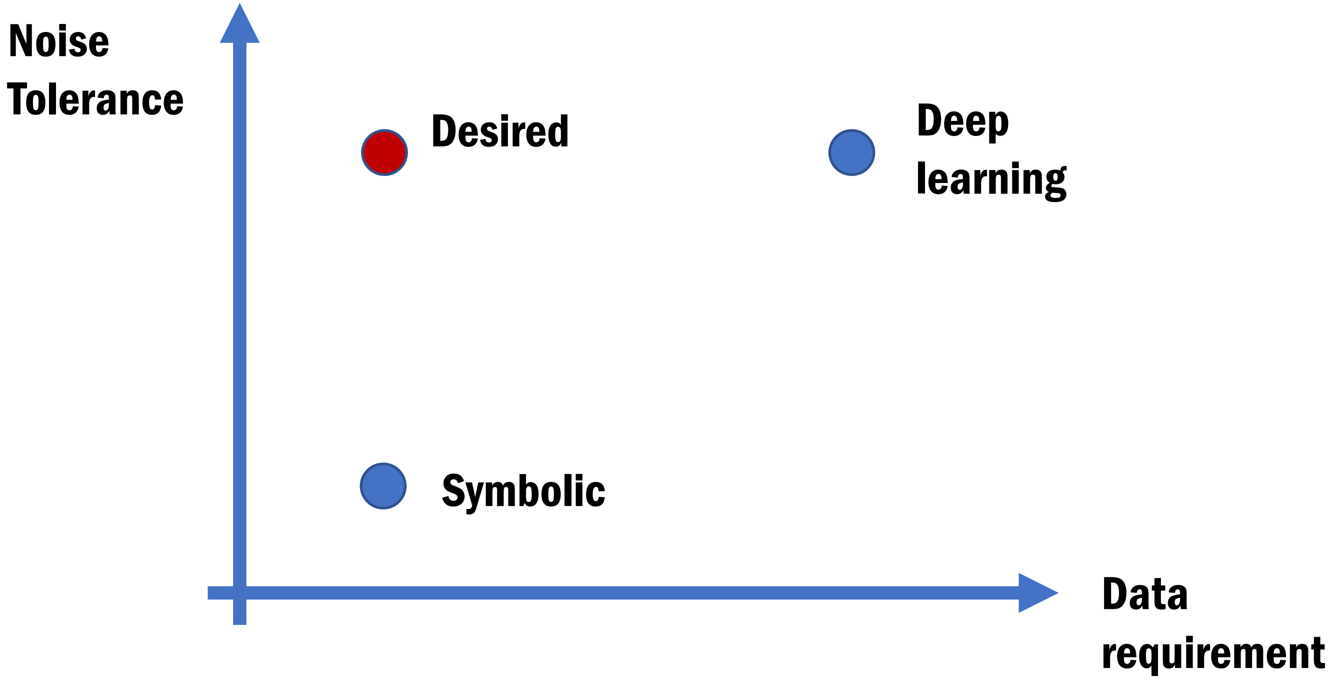
Associate Professor, CSE, College of Computing

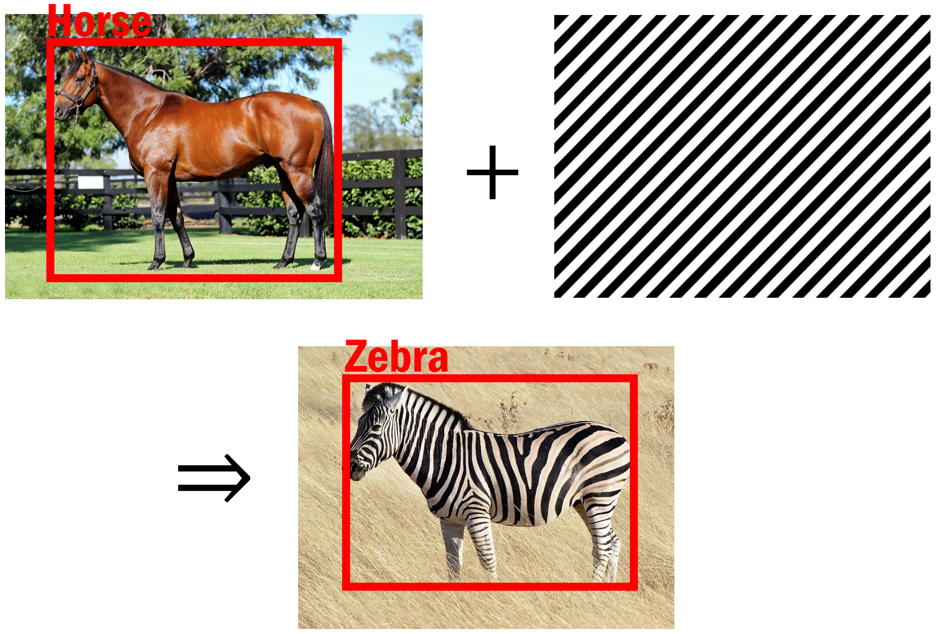
Associate Director, Center for Machine Learning

Georgia Institute of Technology

Learning with small data

- Combine learning and logic reasoning
    - Learning: generalize with noisy input
    - Logic/symbolic reasoning: predict without data





Example of logic inference:
$\text{HorseShape}(c) \wedge \text{StripePattern}(c) \Rightarrow \text{Zebra}(c)$

Graph neural networks (GNN/Structure2vec)

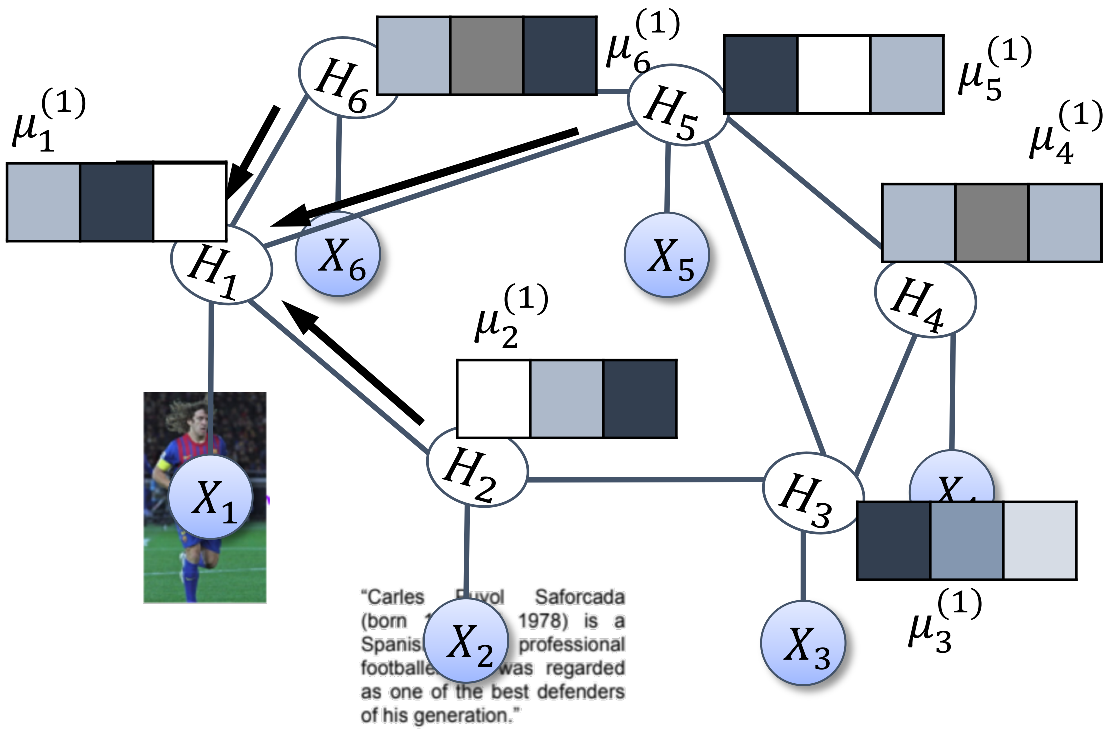
- **Obtain embedding via iterative updates:**

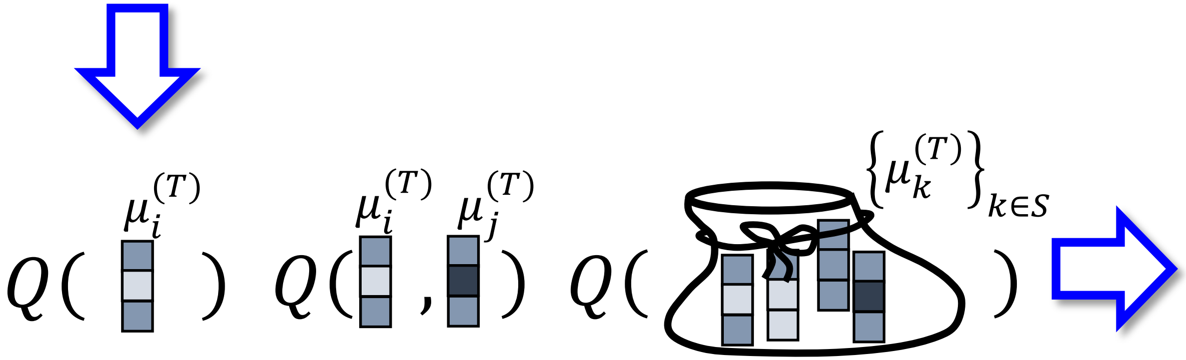
1. **Initialize** $\mu_i^{(0)} = h(W_1 X_i), \forall\ i$

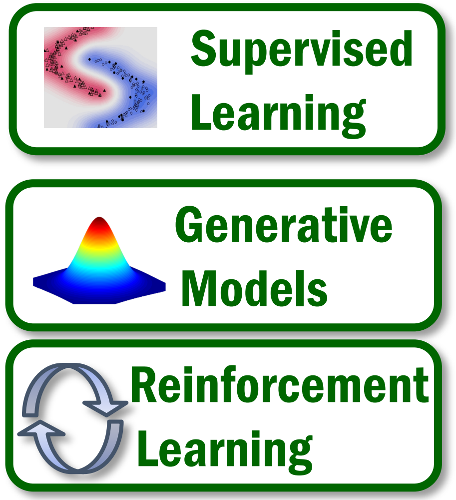
2. **Iterate** T **times**

neural network

$$\boxed{\mu_i^{(t)} \leftarrow h\left(W_1 \mu_i^{(t-1)} + W_2 \sum_{j \in \mathcal{N}(i)} \mu_j^{(t-1)}\right)}$$



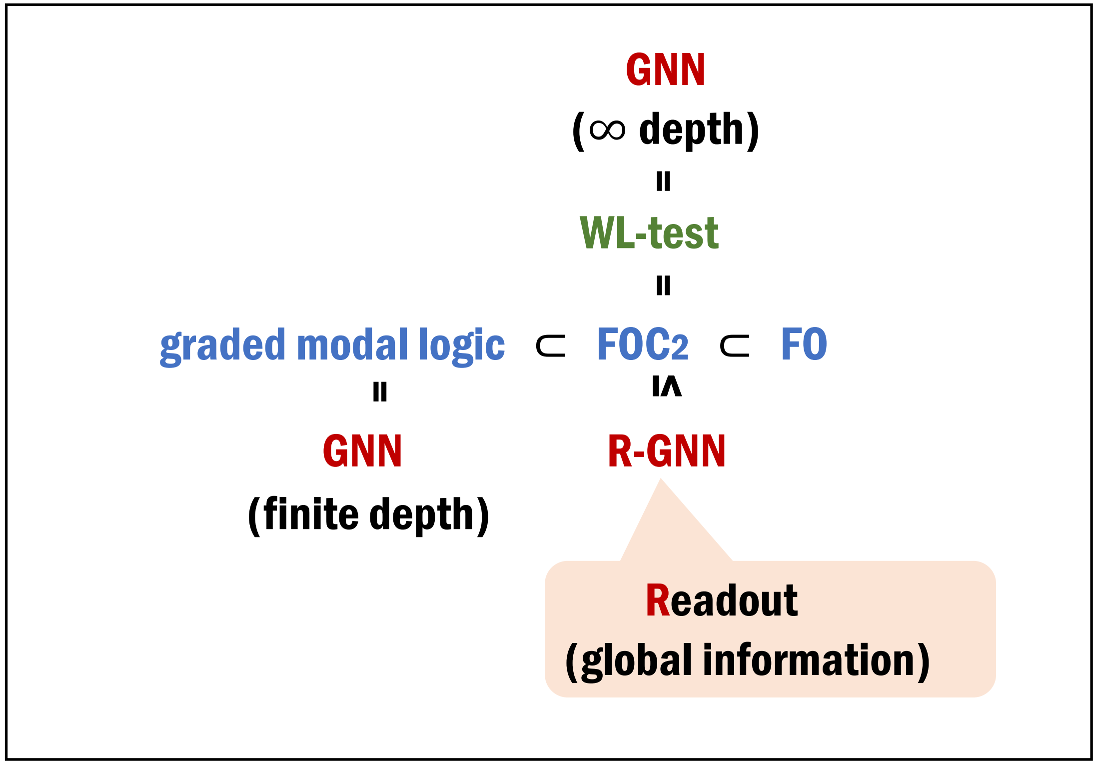




[Dai, et al. ICML 16]

Logical expressiveness of GNN

- **Logical classifiers**
  - classifies each node according to whether it satisfies a formula
  - the formula is expressed in first order predicate logic (FO)



FOC2

- **2 variable**
- **With Counting quantifier,** $\exists^{\geq N}$

$$f(x) \coloneqq \mathrm{Red}(x) \land \exists^{\geq 3} y\ [\neg \mathrm{Edge}(x, y) \land Blue(y)]$$

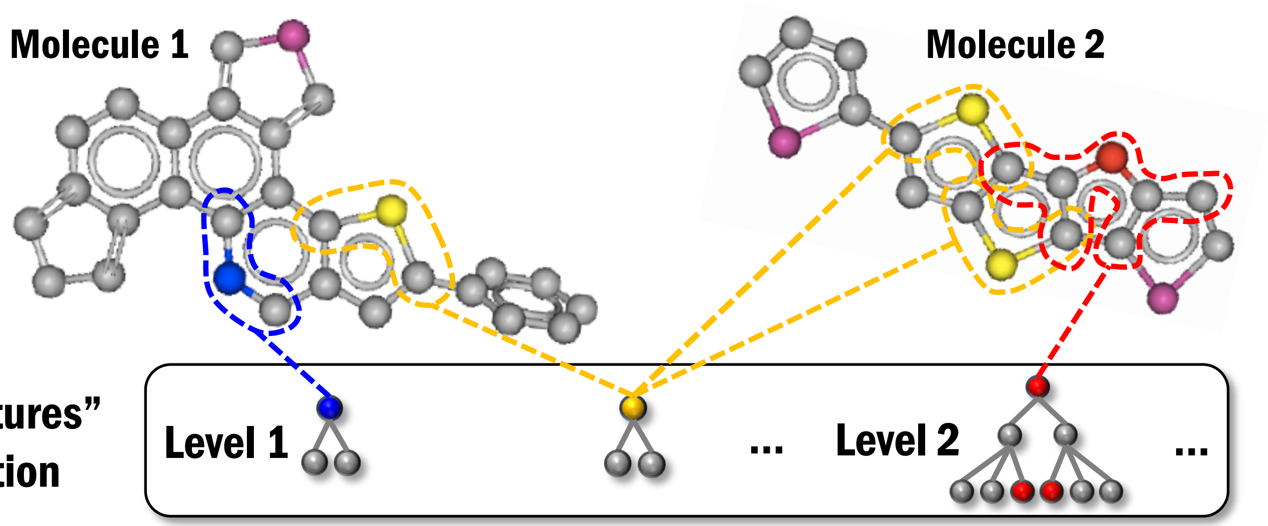
graded modal logic ⊂ FOC2

- **Guarded by** $\mathrm{Edge}(x, y)$.
- **Allow** $\exists^{\geq N} y\ [\mathrm{Edge}(x, y) \land \mathrm{Predicate}(y)]$,
- **but do not allow** $\exists^{\geq N} y\ [\mathrm{Predicate}(y)]$

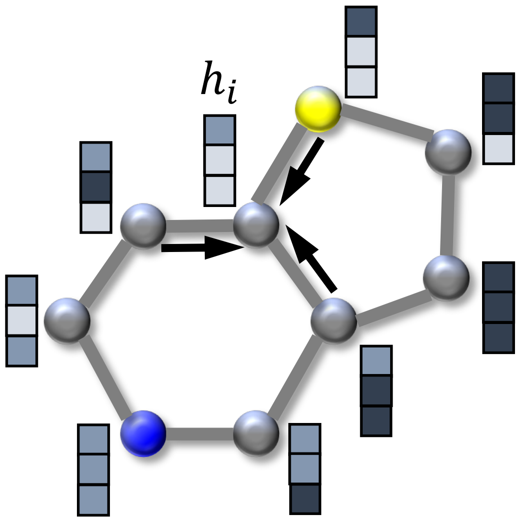
$$f(x) \coloneqq \mathrm{Red}(x) \land \exists^{\geq 3} y\ [\mathrm{Edge}(x, y) \land Blue(y)]$$

[Barcelo, et al. ICLR'20]

Reasoning about subtree structures



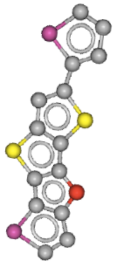
"Bag of structures" representation



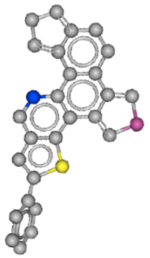
Weisfeiler-Lehman kernel

1. $h_i \leftarrow$ **Hash(node type)**, $\forall\ i$
2. **Iterate** T **times:**
   $h_i \leftarrow$ **Hash(** h_i **+** $\sum_{j \in \mathcal{N}(i)} h_j$ **)**, $\forall\ i$
3. **Aggregate** $\sum_{\forall\ i} h_i$

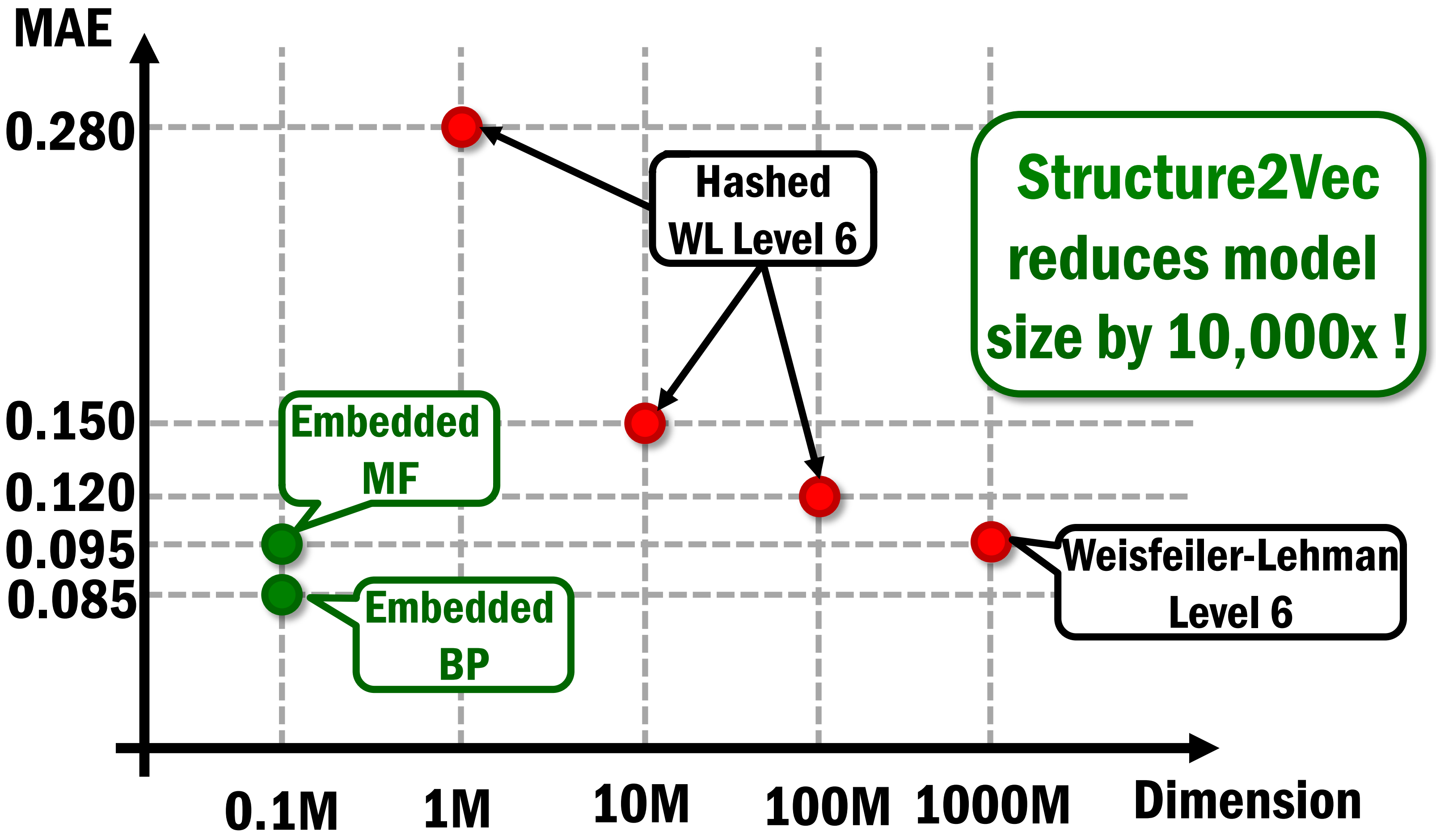
[Shervashidze, et al. JMLR 11]



More compact representation and lower error



Harvard clean energy dataset, 2.3 million organic molecules, predict power conversion efficiency (0 -12 %)



[Dai, et al. ICML'16]

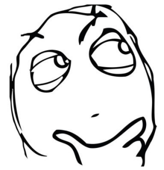
Fraudulent account detection

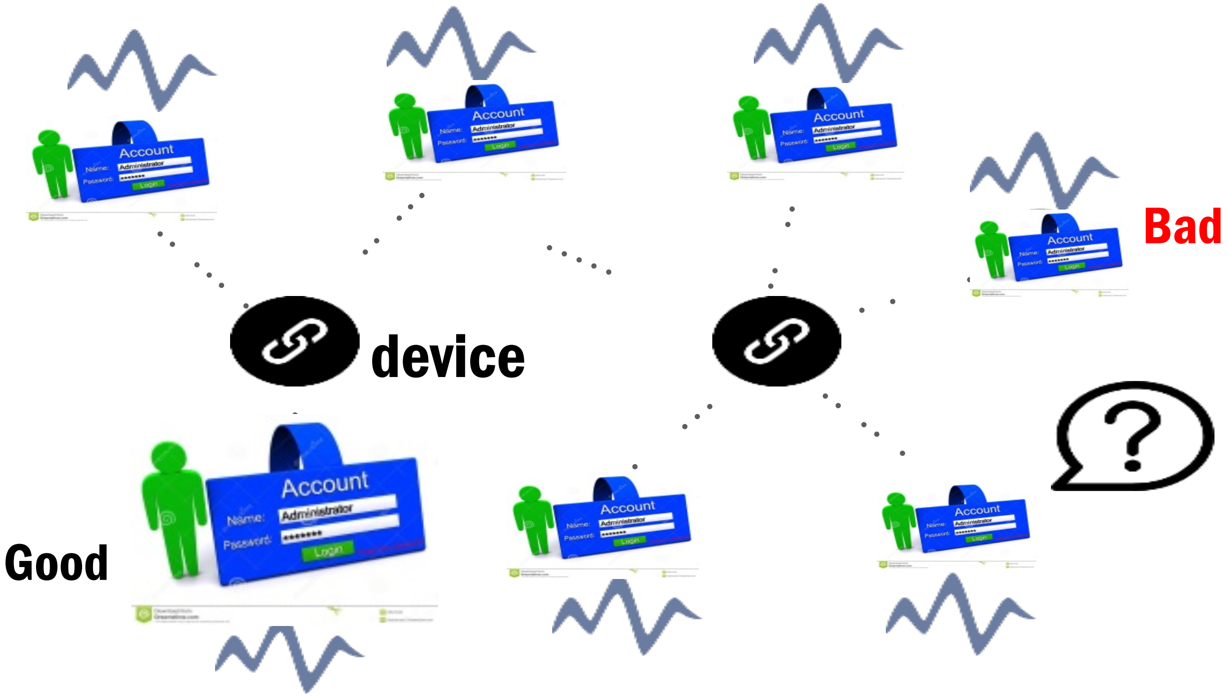
Account – Device Network

Alipay: online payment platform

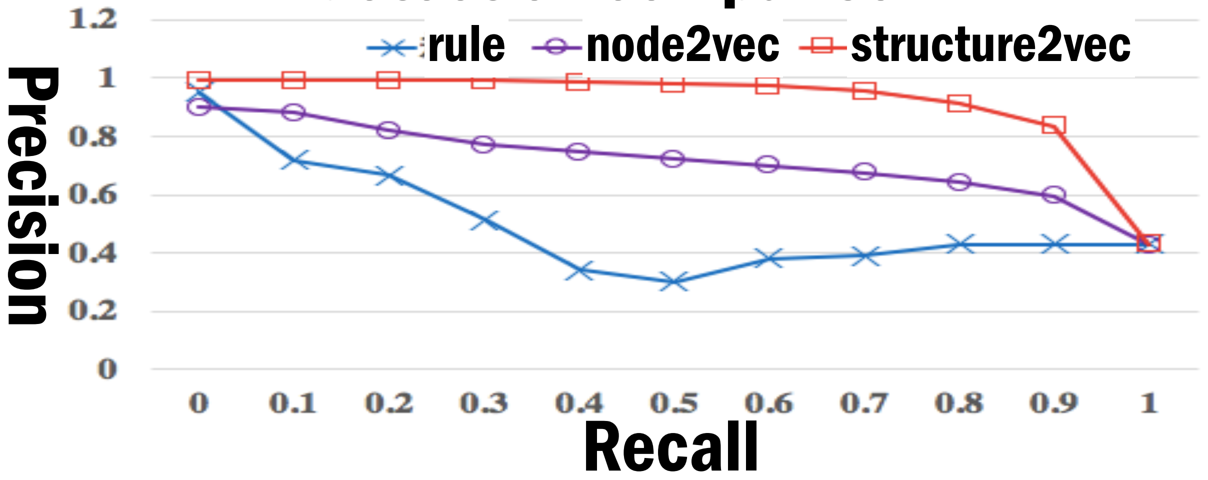
new accounts in a month:
millions of nodes and edges.

Fake account can increase
system level risk





detection comparison



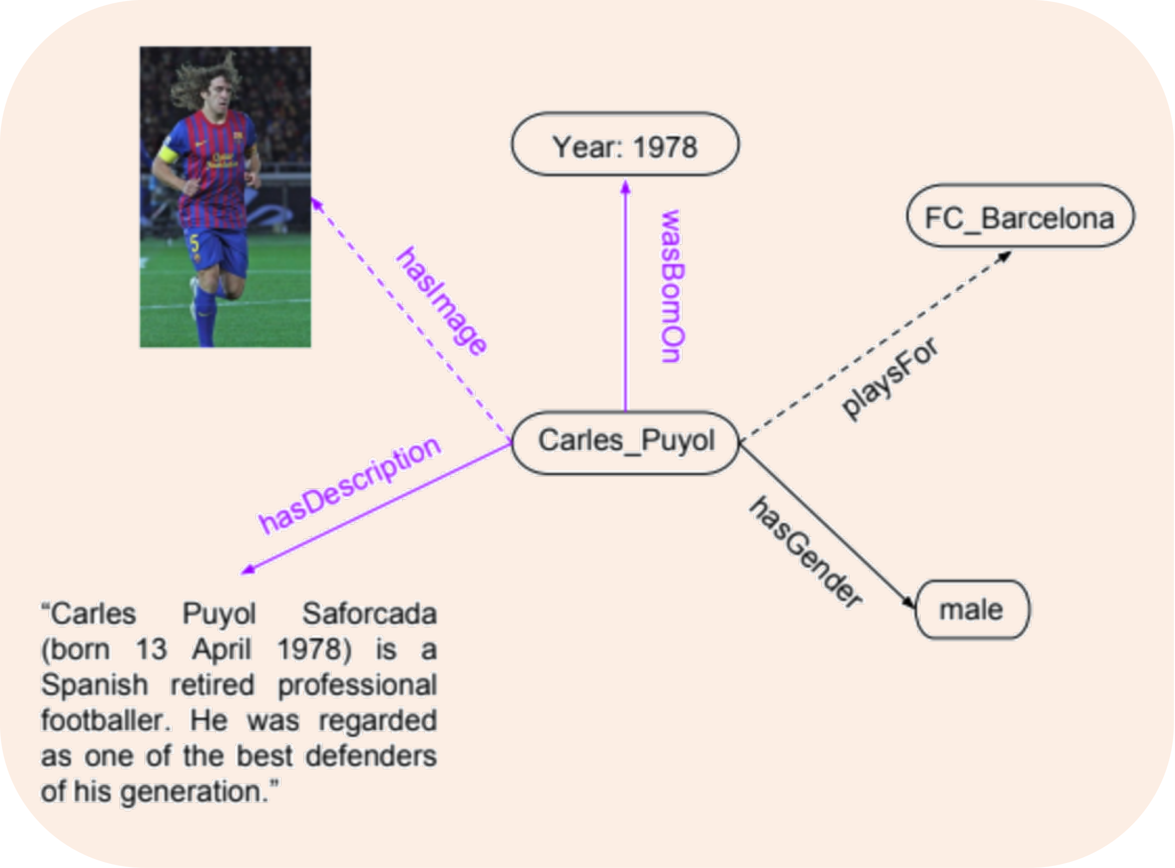
[Liu, et al. CCS'17, CIKM'17, AAAI'19]

Knowledge base reasoning

- Many domains need knowledge bases ...
  - Contain incorrect, incomplete, duplicated records
  - Need link prediction, attribution classification, record de-duplication.

- Many existing rules/logic

- Graph neural networks
  - not explicitly use existing knowledge



UW-CSE dataset details

- **22 relations**
  - **Teach, publish ...**
- **Task goal**
  - **Predict who is whose advisor**
  - **Zero observed facts for query predicates**
- **94 crowd-sourced FOL formulas**

advisedBy(s, p) ⇒ professor(p)

advisedBy(s, p) ⇒ ¬yearsInProgram(s, Year_1)

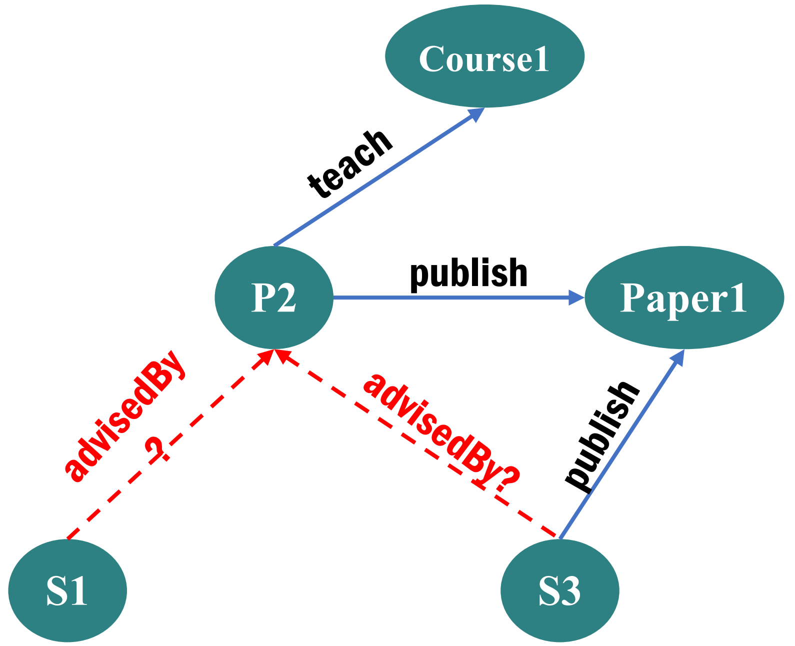
professor(x) ⇒ ¬student(y)

publication(p, x) v publication(p,y) v student(x) v ¬student(y) ⇒ professor(y)

student(x) v ¬advisedBy(x,y) ⇒ tempAdvisedBy(x,y)

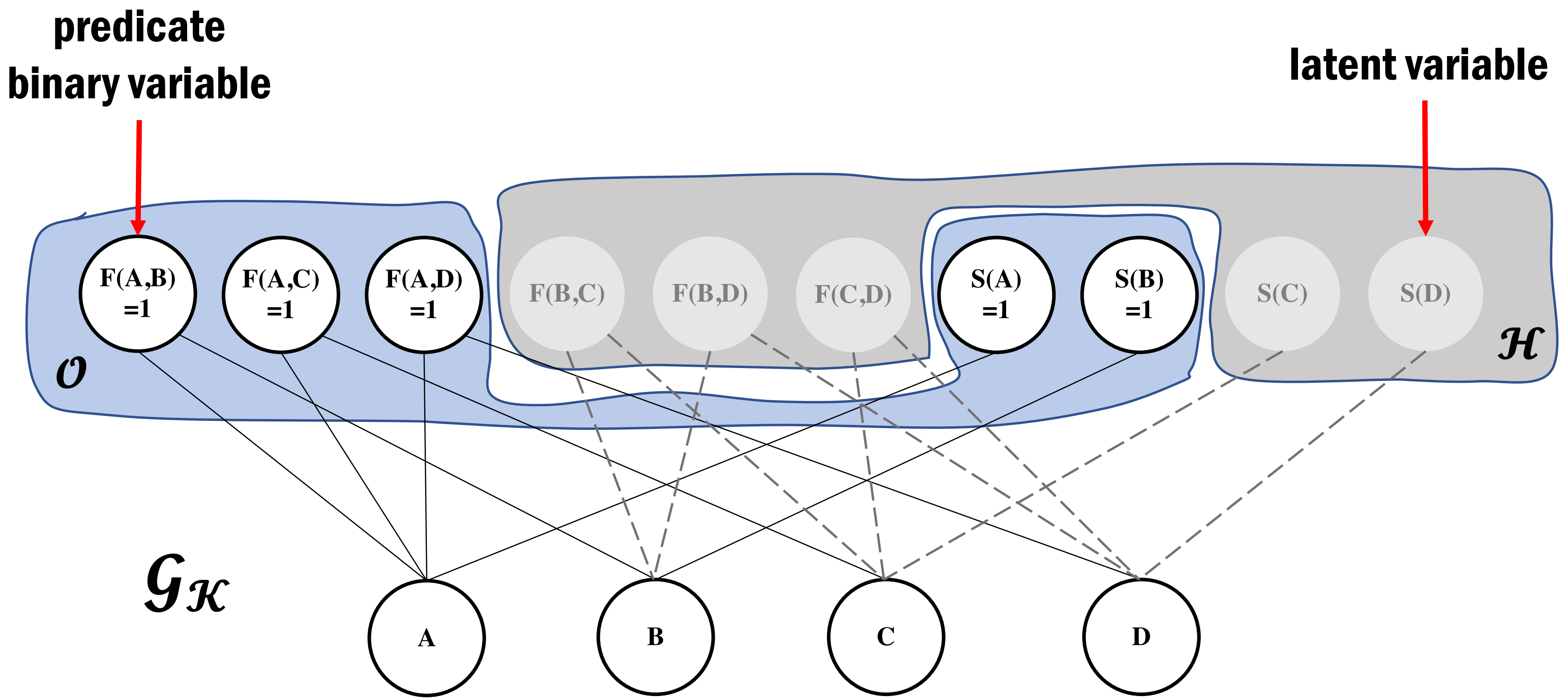
...

Counting	UW-CSE				
	AI	Graphics	Language	Systems	Theory
# entity	300	195	82	277	174
# relation	22	22	22	22	22
# fact	731	449	182	733	465
# query	4K	4K	1K	5K	2K
# ground atom	95K	70K	15K	95K	51K
# ground formula	73M	64M	9M	121M	54M



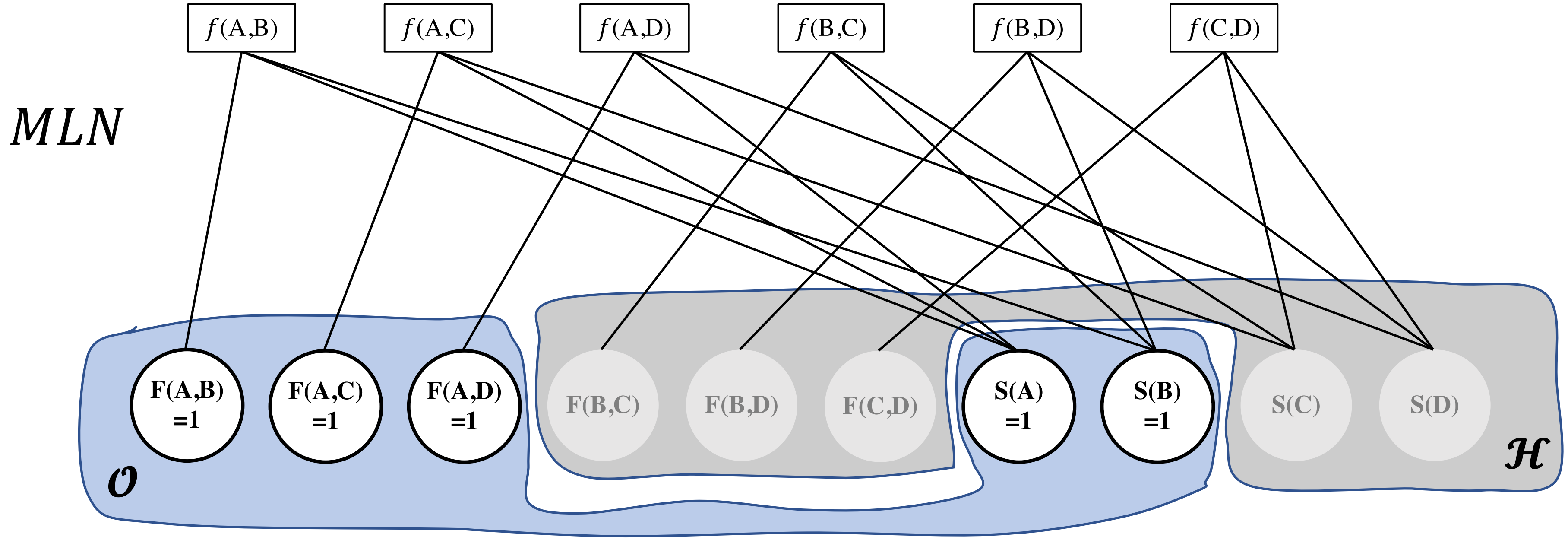
Bipartite graph representation for knowledge base

- **Entity,** $\mathcal{C} = \{A, B, C, D \ldots\}$
- **Predicate (attribute | relation),** $r(\cdot): \mathcal{C} \times \mathcal{C} \times \cdots \mapsto \{0,1\}$
  - **Eg. Smoke(x), Friend(x,x'), Like(x,x')**



Markov logic networks

- Use logic formula $f(\cdot): \mathcal{C}\times\mathcal{C}\times\cdots\times\mathcal{C} \mapsto \{0,1\}$ for potential functions
  - Eg. formula f(x,x'): Friend(x,x') ∧ Smoke(x) ⇒Smoke(x')

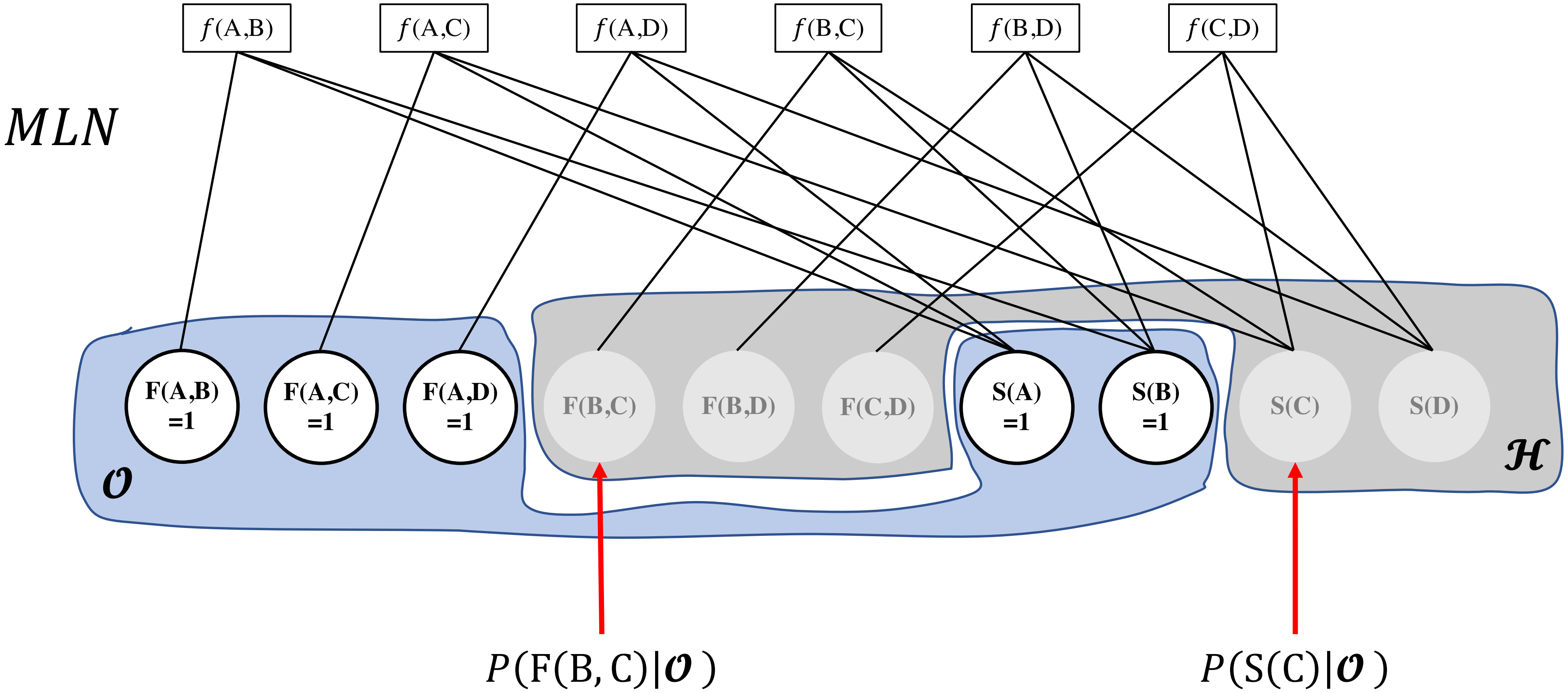


$$P(\mathcal{O},\mathcal{H}) = \frac{1}{Z}\exp\left(\sum_f w_f \sum_{a_f} \phi_f(a_f)\right)$$

w_f: formula weight, eg. $\phi_f(x,x')$: ¬Friend (x,x')∨¬Smoke (x) ∨ Smoke (x')

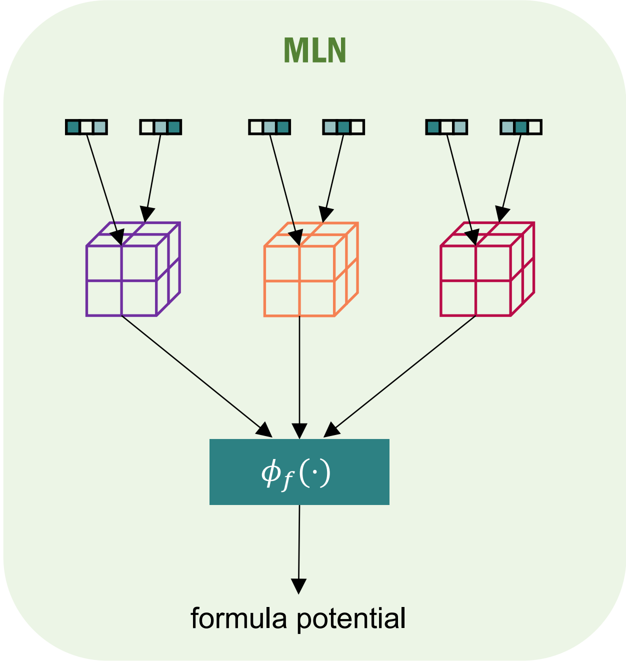
Challenges in inference

- **Large grounded network, $O(n^2)$ in the number n of entities!**
- **Enumerate configuration over $O(n^2)$ binary variables, with $O\left(2^{n^2}\right)$ possibilities.**



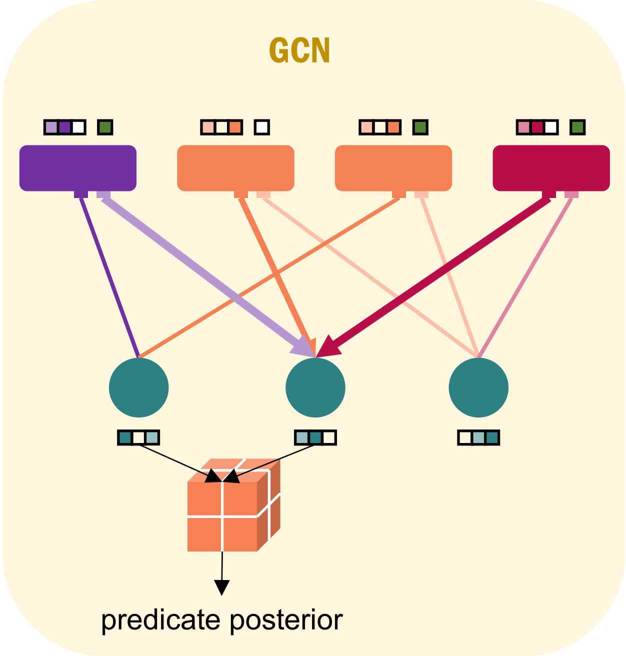
$$Z = \sum_{\text{all possible combination of variable values}} \exp\left(\sum_f w_f \sum_{a_f} \phi_f(a_f)\right)$$

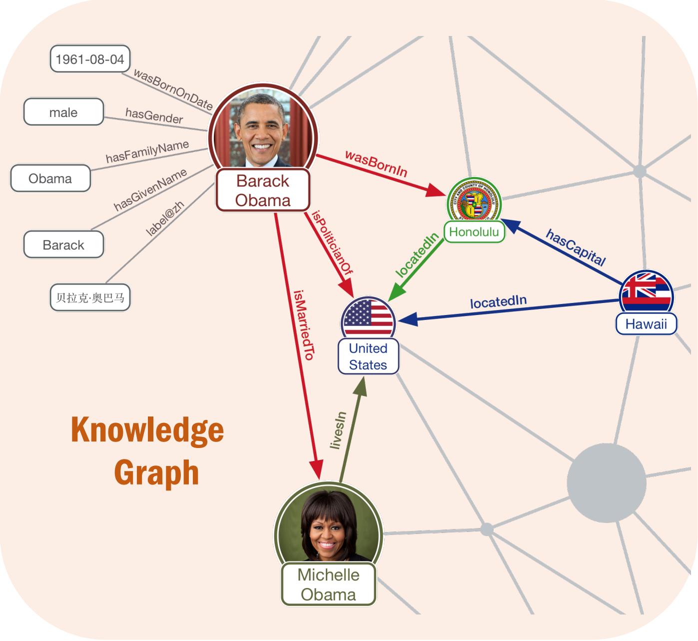
Variational EM



Variational EM

$$\log P(O) \geqslant \mathbb{E}_{Q(\mathcal{H} \mid O)}[\log P(O, \mathcal{H})] - \mathbb{E}_{Q(\mathcal{H} \mid O)}[\log Q(\mathcal{H} \mid O)]$$





Likelihood

$$P(\mathcal{O}, \mathcal{H}) = \frac{1}{Z} \exp\left(\sum_f w_f \sum_{a_f} \phi_f(a_f)\right)$$

Posterior

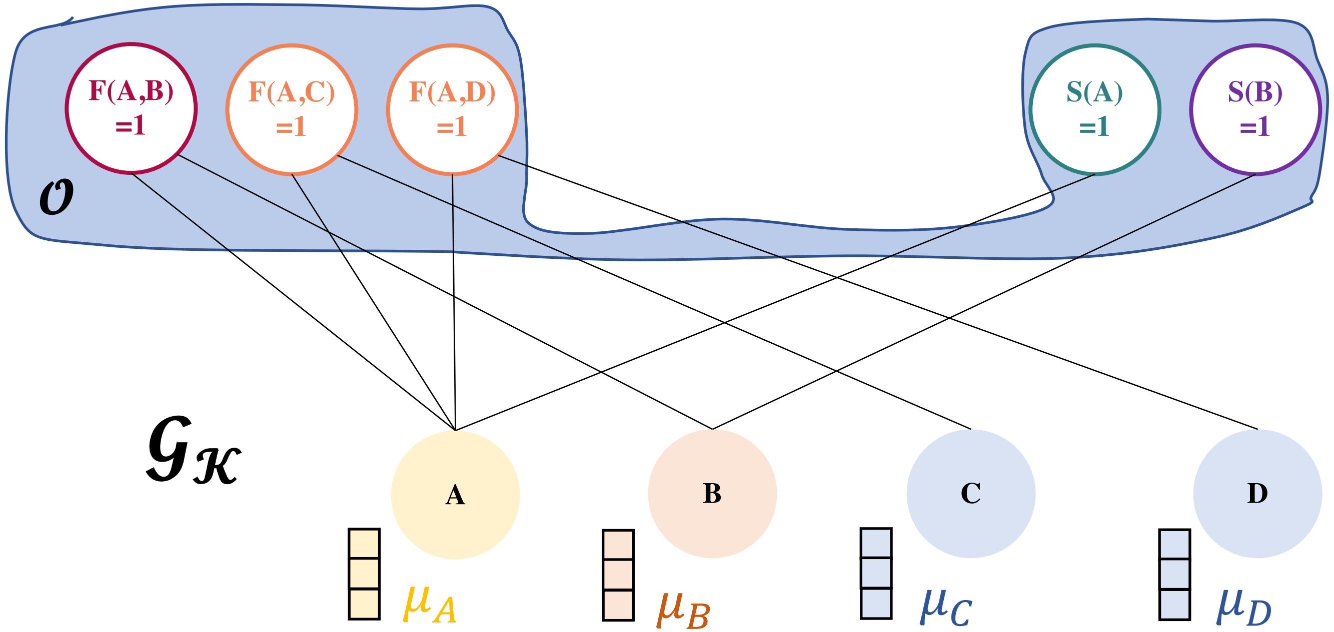
$$Q(\mathcal{H} | \mathcal{O})$$

[Zhang ICLR'20]

ExpressGNN: use GNN in variational inference

- GNN on original KB ($\mathcal{G}_{\mathcal{K}}$) to get embedding $\mu_A, \mu_B, \mu_C, \mu_D$ for entities

$$\mu_A, \mu_B, \mu_C, \mu_D = GNN(\mathcal{G}_{\mathcal{K}}; \theta)$$



graph neural networks

$$\mu_i^{(t)} \leftarrow h\left(W_1 \mu_i^{(t-1)} + W_2 \sum_{j \in \mathcal{N}(i)} \mu_j^{(t-1)}\right)$$

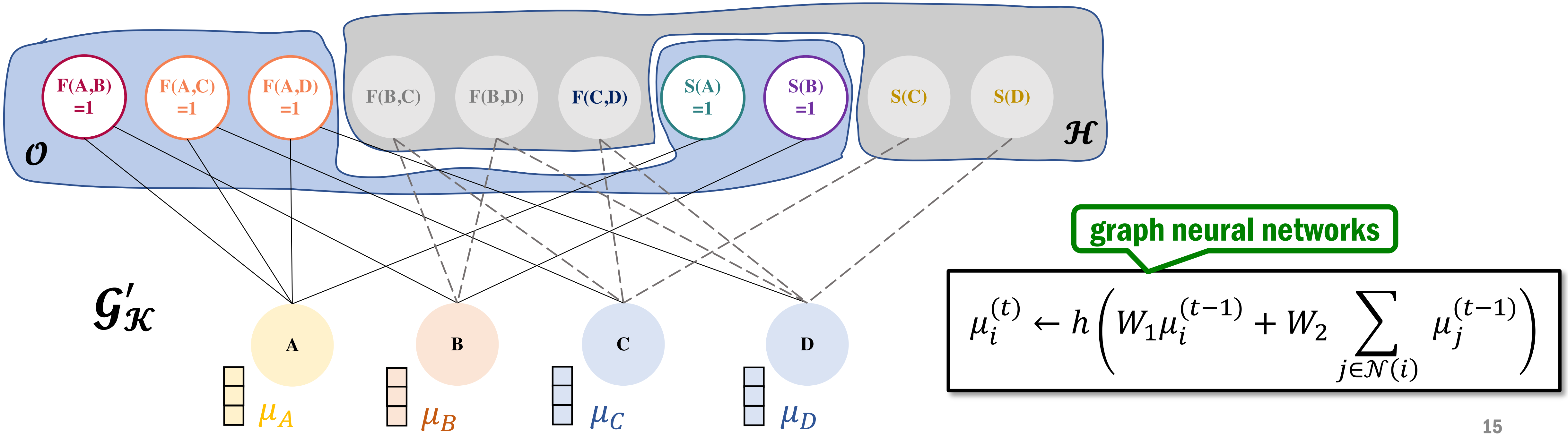
[Zhang ICLR'20]

ExpressGNN: use GNN in variational inference

- **GNN on original KB ($\mathcal{G}_{\mathcal{K}}$) to get embedding $\mu_A, \mu_B, \mu_C, \mu_D$ for entities**

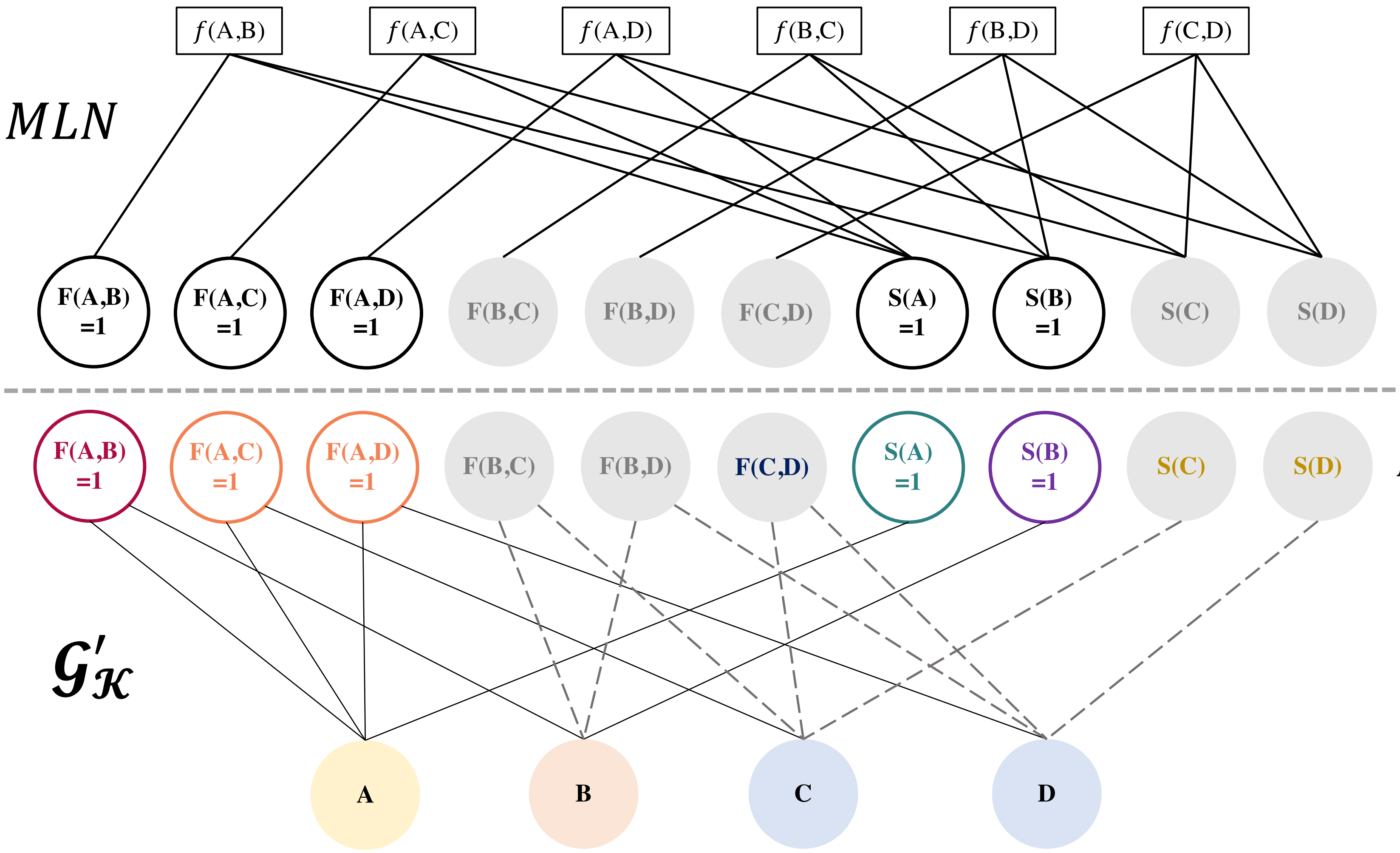
$$\mu_A, \mu_B, \mu_C, \mu_D = GNN(\mathcal{G}_{\mathcal{K}}; \theta)$$

- $Q(\mathrm{F(B,C)} = 1|\mathcal{O}) := \frac{1}{1+\exp(\mu_B^\top \Theta_F\, \mu_C + \omega_B^\top \omega_F)}$, $Q(\mathrm{S(C)} = 1|\mathcal{O}) := \frac{1}{1+\exp(\theta_S^\top [\mu_C, \omega_C])}$



[Zhang ICLR'20]

Is GNN embedding expressive enough?



$$Q(\mathrm{F(B,C)} = 1 | \mathcal{O}) := \frac{1}{1+\exp(\mu_B^\top \Theta_F \, \mu_C + \omega_B^\top \omega_F)}$$

$$F(x,y) \overset{MLN}{\Longleftrightarrow} F(x',y')$$

not true

$$F(x,y) \overset{\mathcal{G}'_{\mathcal{K}}}{\Leftrightarrow} F(x',y')$$

if and only if

$x \overset{\mathcal{G}_{\mathcal{K}}}{\Leftrightarrow} x'$ **and** $y \overset{\mathcal{G}_{\mathcal{K}}}{\Leftrightarrow} y'$

[Zhang ICLR'20]

Cora dataset details (CS paper citations)

- **10 relation types**
  - Author, Title, Venue, HasWordTitle, …

- **Task goal (entity resolution)**
  - De-duplicate citations, authors, and venues
  - Zero observed facts for query predicates

- **46 crowd-sourced FOL formulas**

Author(bc1,a1) v Author(bc2,a2) v SameAuthor(a1,a2) ⇒ SameBib(bc1,bc2)

HasWordAuthor(a1, w) v HasWordAuthor(a2, w) ⇒ SameAuthor(a1, a2)

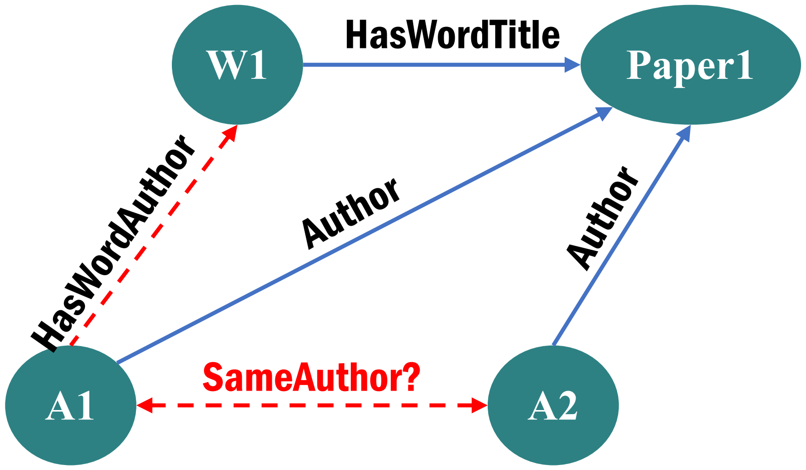
Title(bc1,t1) v Title(bc2,t2) v SameBib(bc1,bc2) ⇒ SameTitle(t1,t2)

SameVenue(v1,v2) v SameVenue(v2,v3) ⇒ SameVenue(v1,v3)

Title(bc1, t1) v Title(bc2, t2) v HasWordTitle(t1, +w) v HasWordTitle(t2, +w) ⇒ SameBib(bc1, bc2)

…

Counting	Cora (average)
# entity	616
# relation	10
# fact	12K
# query	2K
# ground atom	157K
# ground formula	457M



Inference accuracy and time

- **Area under precision-recall curve (AUC-PR)**

Method	UW-CSE					Cora
	AI	Graphics	Language	Systems	Theory	(avg)
MCMC	-	-	-	-	-	-
BP / Lifted BP	0.01	0.01	0.01	0.01	0.01	-
MC-SAT	0.03	0.05	0.06	0.02	0.02	-
HL-MRF	0.06	0.06	0.02	0.04	0.03	-
ExpressGNN	**0.09**	**0.19**	**0.14**	**0.06**	**0.09**	**0.64**

- **Inference wall clock time**

Method	Inference Time (minutes)				
	AI	Graphics	Language	Systems	Theory
MCMC	>24h	>24h	>24h	>24h	>24h
BP	408	352	37	457	190
Lifted BP	321	270	32	525	243
MC-SAT	172	147	14	196	86
HL-MRF	135	132	18	178	72
ExpressGNN-E	**14**	**20**	**5**	**7**	**13**

Large Facebook15K-237 dataset

- # entity: 15K, #ground predicate: 50M, #ground formula: 679B

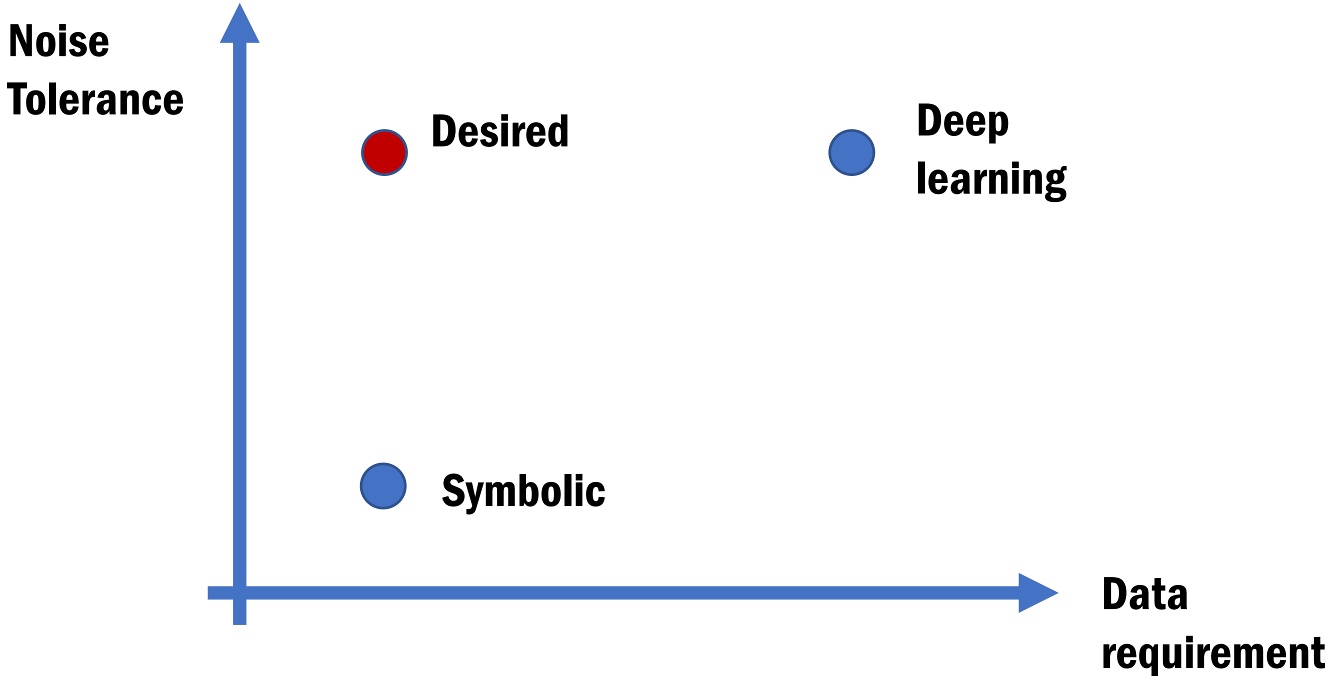
FB15K-237	position(B, A) ∧ position(C, B) ⇒ position(C, A)
	ceremony(B, A) ∧ ceremony(C, B) ⇒ categoryOf(C, A)
	film(B, A) ∧ film(C, B) ⇒ participant(A, C)
	storyBy(A, B) ⇒ participant(A, B)
	adjoins(A, B) ∧ country(B, C) ⇒ serviceLocation(A, C)

- Increase training data for target predicates

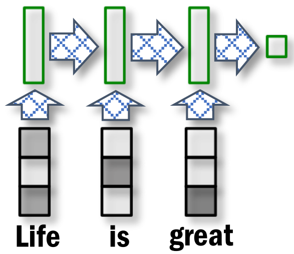
Model	MRR					Hits@10				
	0%	5%	10%	20%	100%	0%	5%	10%	20%	100%
Neural LP	0.01	0.13	0.15	0.16	0.24	1.5	23.2	24.7	26.4	36.2
NTN	0.09	0.10	0.10	0.11	0.13	17.9	19.3	19.1	19.6	23.9
TransE	0.21	0.22	0.22	0.22	0.28	36.2	37.1	37.7	38.0	44.5
ExpressGNN	**0.42**	**0.42**	**0.42**	**0.44**	**0.45**	**53.1**	**53.1**	**53.3**	**55.2**	**57.3**

Learning with small data

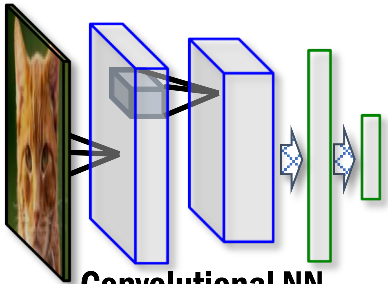
- Combining learning and logic reasoning
  - Learning: generalize with noisy input
  - Logic/symbolic reasoning: predict without data



Learning based approaches



Recurrent NN



Convolutional NN

Symbolic reasoning:

HorseShape(c) ∧ StripePattern(c) ⇒ Zebra(c)

$$\frac{-\hbar^2}{2m}\nabla^2\Psi(\mathrm{r}) + V(r)\Psi(\mathrm{r}) = E\Psi(\mathrm{r})$$

Kinetic Energy + *Potential Energy* = *Total Energy*